

Ten Challenges for Normative Multiagent Systems*

Guido Boella
Dipartimento di Informatica
Università di Torino
Italy
`guido@di.unito.it`

Leendert van der Torre
Computer Science and Communication Research Unit
University of Luxembourg
Luxembourg
`leendert@vandertorre.com`

Harko Verhagen
Department of Computer and Systems Sciences
Stockholm University / KTH
Sweden
`verhagen@dsv.su.se`

September 12, 2008

Abstract

In this paper we discuss the shift from a legal to an interactionist view on normative multiagent systems, examples, and ten new challenges in this more dynamic setting.

1 Towards a more dynamic interactionist view

Traditionally normative systems have been studied in philosophy, sociology, law, and ethics, and during the past two decades they have been studied in deontic logic in computer science (Δ EON). Normative multiagent systems is a research area where the traditional normative systems and Δ EON research fields meet agent research. The proposed solutions to the Δ EON research problems are changing, and solutions based on multiagent systems are increasing. Gradually

*This paper appeared as the editorial [4].

the Δ EON research focus changes from logical relations among norms, to, for example, agent decision making, and to systems in which norms are created and in which agents can play the role of legislators. The eighth conference on Deontic Logic in Computer Science in 2006 in Utrecht, the Netherlands had as special focus “artificial normative systems” [12, 11], and the seventh conference [15, 16] in 2004 in Madeira, Portugal had as special theme “deontic logic and multiagent systems.” Continuing this trend, the third workshop on normative multiagent systems is co-located in Luxembourg in July 2008 with the ninth conference on Deontic Logic in Computer Science [12, 24], which has as special topic “security and trust,” and the fourth workshop will again be a Dagstuhl seminar to be held in March 2009.

The Agentlink Roadmap [17, Fig. 7.1.] observes that norms must be introduced in agent technology in the medium term for infrastructure for open communities, reasoning in open environments and trust and reputation. After four days of discussion, the participants of the second workshop on normative multiagent systems agreed to the following consensus definition:

“A normative multiagent system is a multiagent system organized by means of mechanisms to represent, communicate, distribute, detect, create, modify, and enforce norms, and mechanisms to deliberate about norms and detect norm violation and fulfilment.”

The shift towards a more dynamic interactionist view on normative multiagent systems is reflected in the way this definition builds on its predecessor which emerged at the first workshop on normative multiagent systems held in 2005 as a symposium of the Artificial Intelligence and Simulation of Behaviour convention (AISB) in Hatfield, United Kingdom: “A normative multiagent system is a multiagent system together with normative systems in which agents on the one hand can decide whether to follow the explicitly represented norms, and on the other the normative systems specify how and in which extent the agents can modify the norms” [3]. The emphasis has shifted from representation issues to the mechanisms used by agents to coordinate themselves, and in general to organize the multiagent system. Norms are communicated, for example, since agents in open systems can join a multiagent system whose norms are not known. Norms are distributed among agents, for example, since when new norms emerge the agent could find a new coalition to achieve its goals. Norm violations and norm compliance are detected, for example, since spontaneous emergence norms of among agents implies that norm enforcement cannot be delegated to the multiagent infrastructure.

This shift of interest marks the passage of focus from the more static legalistic view of norms (where power structures are fixed) to the more dynamic interactionist view of norms (where agent interaction is the base for norm related regulation. This ties in to what Strauss [21] called “negotiated order”, Goffmans [13] view on institutions, and Giddens’ [10] structuration theory). The workshop vote on next generation scenarios for normative multiagent systems clearly preferred social scenarios like virtual communities and Second Life

(over 50%) to more classical e-commerce settings where centralized solutions like e-institutions are used (less than 20%).

The legalistic view of normative multiagent systems is a top-down view which considers the normative system as a regulatory instrument to regulate emerging behavior of open systems without enforcing the desired behavior. Agents are often motivated by sanctions to stick to norms, rather than by their sharing of the norms. Even if agents are allowed some freedom to create norms, this freedom is mostly restricted to the possibility for agents to create contracts to regulate the interaction among them.

The interactionist view on normative multiagent systems represents a bottom-up view. In this autonomous individually oriented view norms can be seen, e.g., as regularities of behavior which emerge without any enforcement system because agents conform to them either because their goals happen to coincide, or because they feel themselves as part of the group or because they share the same values of other agents. Sanctions are not always necessary, where by sanctions we mean formal measures towards norm violating agents carried out by agents whose task it is to sanction norm violations, because social blame and spontaneous exclusion of non-conforming agents are often enough. This interactionist view, which has been promoted in the multiagent systems community by Cristiano Castelfranchi [8], becomes essential in applications related to virtual communities. In Second Life, for example, communities emerge in which the behavior of its members show increasing homogeneity.

To put this shift from legal to interactionist view into perspective, we can identify five levels in the development of normative multiagent systems. At level 1 of off-line norm design [20], norms are imposed by the designer and automatically enforced, and agents cannot organize themselves by means of norms. At level 2 of norm representation, norms are explicitly represented, they can be used in agent communication and negotiation, and a simple kind of organizations and institutions can be created. At level 3 of norm manipulation, a legal reality is created in which agents can add and remove norms following the rules of the normative system. Whereas existing normative multiagent systems are still at one of these first three levels of norm autonomy (for an introduction to norm autonomy in multiagent systems, see [25]), multiagent system research is now moving to level 4 of social reality, and is concerned with the ten challenges discussed in Section 3 below. We believe that there is at least one more level to be dealt with in the future. At level 5, the norms create a new moral reality. This goes beyond present studies in machine ethics [1], which is more concerned with agent decision making in the context of norms, which is an issue dealt with at each level of normative multiagent systems, than with creating a new ethics.

Clearly, for each level the development of the normative multiagent system will take a much larger effort than the development of similar systems at lower levels. For example, if norm are explicitly represented (level 2) rather than built into the system (level 1), then the system has to be much more flexible to deal

with the variety of normative systems that may emerge. However, it may be expected that normative multiagent systems realized at higher levels will have a huge effect on social interaction, in particular on the web. In the following sections, we discuss some examples and several research needs that arise in this more dynamic interactionist view on normative multiagent systems.

2 Examples of an interactionist view

We illustrate the more dynamic interactionist viewpoint on normative multiagent systems using virtual communities in virtual reality settings like Second Life. In these virtual communities, human agents interact with artificial agents in a virtual world. The new communication instruments offered by the internet have resulted in the creation of virtual communities of users sharing information, emotions, or hobbies. When the interaction possibilities are multiplied in applications like Second Life or multi-player online games, new scenarios emerge. In particular, given the higher degree of freedom of behavior with respect to the real world, and the unaccountability offered by anonymity, on the one hand, as said above, spontaneous communities emerge showing regularities of behavior. However, to preserve the autonomy of the members of these communities, interactionist mechanisms for regulating behavior are needed. Thus, members of communities should be endowed with tools to make the community norms explicit and communicable to preserve their members' autonomy.

The participants will eventually end up creating their own norms and rules, even if in virtual communities like Second Life and in multi-player games normative infrastructure is imposed by the designers. Sometimes, the rules created by the participants counter the designers' objectives and rules and players start to play in ways unforeseen by the game designers. An example is discussed by Peter Ludlow [18] from Sony's EverQuest. EverQuest is a multiplayer online game where gamers are supposed to fight each other in a world of snakes, dragons, gods, and the Sleeper. Sony intended the Sleeper to be unkillable and gave it extreme high hit points. However, a combined combat of close to 200 players nearly succeeded to kill the 'animal'. Unfortunately, Sony decided to intervene and rescue the monster. Most of the discussion on this example has highlighted the decrease in trust of the game players in Sony, despite the fact that the next day Sony let the game players beat the Sleeper. However, in this paper we would like to highlight what this story tells us about the goals of game players, and its consequences for necessary technology in games. The following quote illustrates the excitement in killing the Sleeper.

A supposedly [player-vs.-player] server banded together 200 people. The chat channels across the server were ablaze, as no less than 5,000 of us listened in, with OMG theyre attempting the Sleeper! Good luck d00dz! Everyone clustered near their screens, sharing the thrill of the fight, the nobility of the attempt and the courage of those brave 200. Play slowed to a crawl on every server as whispers turned

to shouts, as naysayers predicted, It can't be done or It will drop a rusty level 1 sword and most of us just held our breath, silently urging them forward. Rumors abounded: If they win, the whole EQ world stops and you get the text from the end of Wizardry 1, or If they win, the president of Sony will log on and congratulate them. With thousands watching and waiting, the Sleepers health inched ever downward.

...

[EverQuest player] Ghenwivar writes, On Monday, November 17th, in the most amazing and exciting battle ever, [EverQuest guilds] Ascending Dawn, Wudan and Magus Imperialis Magicus defeated Kerafyrm, also known as The Sleeper, for the first time ever on an EverQuest server. The fight lasted approximately three hours and about 170180 players from [EverQuest server] Rallos Zeks top three guilds were involved. Hats off to everyone who made this possible and put aside their differences in order to accomplish the impossible. Congratulations RZ!!!!" [18]

Normative multiagent systems study multiagent technology to support the emergent cooperation in online multi-player games like EverQuest [2]. The example illustrates that the game had been so well wrought that a real coalition of communities of players had formed, one that was able to set aside the differences between the communities, at least for a night, in pursuit of a common goal. This was not intended or foreseen by Sony, and getting two hundred people to focus on accomplishing the same task is a challenge.

"Why, you might ask, would anyone waste four hours of their life doing this? Because a game said it couldn't be done.

This is like the Quake freaks that fire their rocket launchers at their own feet to propel themselves up so they can jump straight to the exit and skip 90% of the level and finish in 2 seconds. Someone probably told them they couldn't finish in less than a minute.

Games are about challenges, about hurdles or puzzles or fights overcome. To some players, the biggest hurdle or challenge is how to do what you (the designer) said couldn't happen. If you are making a game, accept this." [18]

A typical problem in virtual communities is caused by the ease in which new participants can enter the community, known as "newbies". The virtual communities should be able to defend itself from dangerous new players, and normative systems are a way to pose virtual gates to such communities. "Griefers would also maintain numerous alts that were sent out into greater Alphaville in attempts to scam and disrupt other houses. Because alts were usually abandoned soon after they had been created, they appeared to others as new characters, and this had the effect of making many players highly suspicious of newbies, and

of generating virtual gated communities in response.” [18] However, a virtual space should be able to deal with honest new participants. It has been noted that existing communities establish practices which tend to exclude newly entered participant in the virtual space: “Processes of norm building were visible, resulting in patterns of established users versus outsiders; new bonds were created, and users experience an appropriation of this newly created virtual public space: parts of the Digital City were ‘taken over’ by active established users who behaved as a closed community and were perceived accordingly by the outsiders.” [23]

As illustrated by the “newbies” example, there are some aspects in which normative systems for virtual communities are more challenging than traditional regulations. For example, the construction of autonomous virtual communities cannot ground itself on an external legal system - apart from most serious cases like frauds going beyond the virtual environment - as in e-commerce applications that ground the validity of online contracts on the relevant human regulations. Consequently, these normative systems should be developed separately, in the same way as different national systems are created independently. Another issue is related to the possibility to augment actions in virtual scenarios: in these scenarios characters can be created with their own behavior that have more abilities than humans in the real world (e.g., flying, walking through wall), objects nor existing in reality, and even places. Moreover, the abilities of characters are not only related to the ones of their players: e.g., an avatar in Second Life entering a dancing room can acquire new dancing abilities which it did not have before and will lose afterwards. Thus the autonomy of characters assumes new dimensions.

3 Ten research challenges for the interactionist view

For the ten challenges posed by the interactionist viewpoint, we take the perspective from an agent programmer, and consider which kinds of tools like programming primitives, infrastructures, protocols, and mechanisms she needs to deal with norms in the example scenario. Similar needs exist at the requirements analysis level, or the design level, but we have chosen for the programming level since it makes the discussion more concrete, and this level is often ignored when norms are discussed. The list is not exhaustive, and there is some overlap between the challenges. Our aim is to illustrate the range of topics which have to be studied, and we therefore do not attempt to be complete.

Challenge 1 *Tools for agents supporting communities in their task of recognizing, creating, and communicating norms to agents.*

Even if social norms emerge informally, e.g., when a community becomes more complex and more open, an explicit representation of norms becomes necessary. There are still numerous philosophical problems for the representation

of norms, see, for example, [14]. However, the new problem is the role of the agents and humans involved in the interaction with the multiagent system.

Challenge 2 *Tools for agents to simplify normative systems, recognize when norms have become redundant, and to remove norms.*

Challenge 2 is the counterpart of Challenge 1, because the natural tendency of overregulation creates the need for a counterbalance. Since all norms come with a cost, for example to process them, to communicate them, to maintain them, or to enforce them, norms should only be introduced when they are really needed, and they should be removed as soon as they are no longer needed. For example, when the number of violations is increasing, this is typically a case where norms must be changed or removed, rather than where norm enforcement has to be increased.

Challenge 3 *Tools for agents to enforce norms.*

If we allow communities of agents to create their own normative multiagent systems, then the issue of how to enforce the norms arises. In case a centralized approach is needed, the infrastructure should support the enforcement of norms created by the communities. In a distributed approach, roles should be defined for agents in charge of monitoring and sanctioning. The virtual environment can offer new opportunities for norm enforcement not found in the usual environments. For example, evidence about agent behaviors can be collected via the logfiles of the system.

Challenge 4 *Tools for agents to preserve their autonomy.*

Challenge 4 is the counterpart of Challenge 3, because there is a natural tendency to enforce norms by regimenting them into the system. The danger highlighted by Castelfranchi [9] is related to the “formalization” of the informal. Norms have the nature of general directives which cannot cover all cases nor avoid all conflicts with other norms. Thus, normative multiagent systems need to preserve the autonomy of agents regarding the making of decisions about norm compliance and norm violation. Agents in charge of monitoring and enforcing norms should be flexible enough to preserve the autonomy of the “norm subject” agents with respect to norm violations, for instance in circumstances that differ from the circumstances which the norms have been defined to preserve and where norm compliance is not advantageous for the normative multiagent system.

Challenge 5 *Tools for agents to construct organizations.*

As the example about EverQuest example shows, cooperation among the participants of virtual reality can result in coalitions which can achieve results which go beyond the ones reachable by their members. This is of great interest for participants in virtual reality, also because Second Life is becoming a place

where business takes place. Thus, participants should be given some facilities and tools which allow the construction and management of organizations to achieve their goals. Note that in the real world such mechanisms exist, first of all the laws which allow the creation of organizations and attribute the responsibilities to different entities. E-institutions as proposed in multiagent systems can be a starting point, but they are often too flat - i.e., not hierarchically organized - and they usually do not support the dynamics of the underlying normative systems by allowing the creation of new norms.

Challenge 6 *Tools for agents to create intermediate concepts and normative ontology, for example to decide about normative gaps.*

In real institutions norms have a fuzzy character in the sense that they are not able to cover all possible situations. In particular because new situations can arise, e.g., due to technological advancement (for instance: is a digital signature the same as an handwritten signature?) This problem increases exponentially in virtual worlds where all kind of new behaviors and objects can be defined. The solution in real normative system is to endow some agents with powers to decide whether a new concept is subsumed by another one. The role of agents in the logical reasoning of a normative system is something which is still missing in the state of the art of the field.

Challenge 7 *Tools for agents to decide about norm conflicts.*

This challenge is related to Challenge 6 since norms do not cover all possible cases and conflicts between norms are possible. Thus agents need a mechanism to take decisions in situations of conflicting norms. The mechanism cannot always be automated, for example because the degree of freedom in virtual world to create new behaviors and objects norms may become underspecified. Thus, the problem is to define normative systems, where, like in human normative systems, roles are defined and role keepers are empowered to take decision when automated reasoning alone is not enough. At some point, the view of the normative system as a self contained logical system is not viable anymore.

Challenge 8 *Tools for agents to voluntarily give up some norm autonomy by allowing automated norm processing in agent acting and decision making.*

In many examples, the autonomy of the agent must be adjusted to the context. In general avatars are graphical representations of users of a system and can be seen as interface agents. Avatars living in Second life are interface agents for human players but also increasingly for autonomous agents. Consider the example above, where new abilities like dancing are automatically added to the avatar. Moreover, even if now prohibited, autonomous agents should be allowed to on the player's behalf cope with events that occur when the player is not online. It is possible to envisage a scenario where avatars are partially programmed to take autonomous decisions when the player is off-line. Among these decisions is whether to comply with norms of the community the avatar is acting into.

Note that these mechanisms are useful not only when the avatar is acting autonomously on behalf of its off-line owner, but also during the activity of the player. In real life norms are often violated just by distraction or ignorance or by lack of resources and the violator does not gain anything by its deviant behavior. The same will eventually happen in virtual worlds, especially when norms to be respected will not be necessarily intuitive or similar to the ones of real world. In these cases, the decision to conform to norms can be left to the avatar and the player can be relieved from this task. E.g., consider the case of communities where nudity of avatars is prohibited. The player could simply leave to its avatar the burden to conform to the norms by automatically disabling actions which are deviant with respect to the norms.

Challenge 9 *Tools for conviviality.*

Since scenarios like Second life are aiming at people having pleasant social interactions, and norms may interfere with the goals of the players, the impact of norms on this dimension must be considered. Norms should not constrain the freedom of participants too much and allow to avoid unpleasant behavior from other agents, but there is also a more subtle effect to be considered. Social interaction is regulated by social conventions, which can be modeled as a sort of institution. Part of the fun of “living” in Second life, like when participating in a carnival or when embodying a character of a drama depends - according to Taylor [22, 7] who calls this effect “conviviality” - is the temporary displacement with respect to the usual norms of social life. In particular, in the sense that in social relations the player acquires new social powers which he does not have in his first life.

The tools for conviviality should study social dependencies among players and indicate how these dependencies can be made less unbalanced by attributing more social powers to some players. Note that, as in the example about automatic learning of dancing abilities in Section 2, adding social powers in a virtual reality can take a more extended sense, since in the real world physical abilities cannot be added. Tools for conviviality should also facilitate the introduction of new participants in a virtual community by addressing the “newbies” problems.

Challenge 10 *Tools for legal responsibility of the agents and their principals.*

Nowadays, agents become subjects of human legislation. For example, it is debated if agents have responsibilities beyond the ones attributed to their owner, or if agents can be really attributed mental states which are to be taken into account in the attribution of responsibilities. However, in scenarios like Second life, new questions arise. Participants accept the rules of the game and they should be made aware whether following the rules of some communities leads to infringement of real legislations.

References

- [1] M. Anderson and S. Leigh Anderson. Machine ethics: Creating an ethical intelligent agent. *AI Magazine*, 28(4):15–26, 2007.
- [2] G. Boella, P. Caire, and L. van der Torre. Norm negotiation in online multi-player games. *Knowledge and Information Systems*, To appear.
- [3] G. Boella, L. van der Torre, and H. Verhagen. Introduction to normative multiagent systems. *Computation and Mathematical Organizational Theory, special issue on normative multiagent systems*, 12(2-3):71–79, 2006.
- [4] G. Boella, L. van der Torre, and H. Verhagen. Introduction to the special issue on normative multiagent systems. *Autonomous Agents and Multiagent Systems*, 2008.
- [5] G. Boella, L. van der Torre, and H. Verhagen, editors. *Normative Multiagent Systems*, Dagstuhl Seminar Proceedings 07122, 2007.
- [6] J. Broersen, M. Dastani, J. Hulstijn, and L. van der Torre. Goal generation in the BOID architecture. *Cognitive Science Quarterly*, 2(3-4):428–447, 2002.
- [7] P. Caire, S. Villata, G. Boella, and L. van der Torre. Conviviality masks in multiagent systems. In *Proceedings of the Seventh International Conference on Autonomous Agents and Multiagent Systems (AAMAS'08)*, 2008.
- [8] C. Castelfranchi. Modeling social action for AI agents. *Artificial Intelligence*, 103(1-2):157–182, 1998.
- [9] C. Castelfranchi. Formalising the informal? Dynamic social order, bottom-up social control, and spontaneous normative relations. *Journal of Applied Logic*, 1(1-2):47–92, 2003.
- [10] A. Giddens. *The Constitution of Society*. University of California Press, 1984.
- [11] L. Goble and J.J. Ch. Meyer, editors. *Deontic Logic and Artificial Normative Systems, 8th International Workshop on Deontic Logic in Computer Science, DEON 2006, Utrecht, The Netherlands, July 12-14, 2006, Proceedings*, volume 4048 of *Lecture Notes in Computer Science*. Springer, 2006.
- [12] L. Goble and J.J. Ch. Meyer. Revised versions of papers presented in the proceeding of the eighth international workshop on deontic logic in computer science (DEON06). *Journal of Applied Logic*, in press.
- [13] E. Goffman. *The Presentation of Self in Everyday Life*. Doubleday, 1959.

- [14] J. Hansen, G. Pigozzi, and L. van der Torre. Ten philosophical problems in deontic logic. In G. Boella, L. van der Torre, and H. Verhagen, editors, *Normative Multi-agent Systems*, volume 07122 of *Dagstuhl Seminar Proceedings*. Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany, 2007.
- [15] A. Lomuscio and D. Nute, editors. *Deontic Logic in Computer Science, 7th International Workshop on Deontic Logic in Computer Science, DEON 2004, Madeira, Portugal, May 26-28, 2004. Proceedings*, volume 3065 of *Lecture Notes in Computer Science*. Springer, 2004.
- [16] A. Lomuscio and D. Nute. Revised versions of papers presented in the proceeding of the seventh international workshop on deontic logic in computer science (DEON04). *Journal of Applied Logic*, 3(3-4), 2005.
- [17] M. Luck, P. McBurney, and C. Preist. *Agent Technology: Enabling Next Generation Computing (A Roadmap for Agent Based Computing)*. AgentLink, 2003.
- [18] P. Ludlow and M. Wallace. *The Second Life Herald*. MIT Press, Cambridge (MA), 2007.
- [19] J.R. Searle. *The Construction of Social Reality*. The Free Press, New York, 1995.
- [20] Y. Shoham and M. Tennenholtz. On social laws for artificial agent societies: Off-line design. *Artificial Intelligence*, 73(1-2):231–252, 1995.
- [21] A. Strauss. *Negotiations: Varieties, Contexts, Processes and Social Order*. San Francisco, Jossey-Bass, 1978.
- [22] M. Taylor. Oh no it isn't: Audience participation and community identity. *Trans, Internet journal for cultural sciences*, 1(15), 2004.
- [23] P. van den Besselaar. E-community versus e-commerce: The rise and decline of the Amsterdam digital city. *AI and Society*, 15(3):280–288, 2001.
- [24] R. van der Meyden and L. van der Torre, editors. *Deontic Logic in Computer Science, 9th International Conference on Deontic Logic in Computer Science, DEON 2008, Luxembourg, July 16-18, 2008, Proceedings*, LNCS, Berlin, in press. Springer.
- [25] H. Verhagen. *Norm Autonomous Agents*. PhD thesis, Department of System and Computer Sciences, The Royal Institute of Technology and Stockholm University, Sweden, 2000.